8-2004

Identification of E. coli Sources in Conesus Lake Sub-watersheds Using BOX A1R- Derived Genetic Fingerprints

Jason Somarelli
The College at Brockport, jsoma002@fiu.edu

Identification of *E. coli* Sources in Conesus Lake Sub-watersheds Using BOX A1R-Derived Genetic Fingerprints

A Thesis Presented to the Faculty of the Department of Environmental Science and Biology of the State University of New York College at Brockport in Fulfillment for the Degree of Master of Science

By:

Jason Somarelli

August 2004

THESIS DEFENSE

APPROVED	NOT APPROVED	MASTER'S DEGREE ADVISORY COMMITTEE	
✓		Joseph Makarewicz	13 Jan 05
		Major Advisor	Date
✓		Robert D. Simon	1/13/05
		Committee Member	Date
✓		Roy A Li	1/13/05
		Committee Member	Date
Chairman, Graduate Committee		Chairman, Dept. of Biological Sciences	1/13/05

Acknowledgements

Thanks to Dr. Joseph Makarewicz, Dr. Rey Sia and Dr. Robert Simon for guidance and inspiration throughout the project. Thanks to Ted Lewis for his limitless patience, intelligence and willingness to help. Thanks also to Dan White, Hilary Richardson, Peter D'Aiuto and the entire Biology graduate student body at SUNY Brockport for support, understanding and encouragement during the difficult and frustrating times of this process. Thanks to my parents, Nicholas and Rhonda Somarelli, and my entire family, without whom I would have neither the aspiration nor the ability to pursue my dreams. And finally, thanks to my other family, the friends in my life who tirelessly keep things in perspective.

Table of Contents

List of Tables

Page #

List of Figures

Page #

Abstract

In the Conesus Lake watershed, best management practices (BMP's) were implemented on farms in the upper watershed and the resultant change in bacterial water quality was assessed using both quantification techniques and Rep-PCR molecular tools during events and non-events. Genetic fingerprints of *Escherichia coli* isolates from a library of known source isolates (n=123) were compared to *E. coli* of unknown origin obtained from stream water samples. The genetic library consisted of *E. coli* sources from cattle, humans, geese and deer. Fecal samples were collected aseptically and *E. coli* was isolated from each of these sources. Genetic fingerprints were obtained from each of the known sources using Rep-PCR. Fingerprints from unknown sources were compared with those in the library using computer based image analysis. These techniques were used to identify sources of *E. coli* in the watershed and accurately demonstrate the effectiveness of the best management practices in the watershed. Results found *E. coli* levels that were significantly higher than that established by EPA for recreational waters (200CFU/100ml) on several occasions throughout the year, especially during event periods. In addition, the sources of *E. coli* in the watershed were identified.

Introduction

Escherichia coli (*E. coli*) and other microbial contaminants are a major concern to the general public, and especially to recreational users of our nation's waterways. Elevated levels of *E. coli* often indicate the presence of other pathogenic bacteria and viruses such as *Salmonella spp.* and the Hepatitis A virus, both of which can cause severe gastroenteritis (US Environmental Protection Agency 1986). In addition, *E. coli* strain O157:H7 can cause bloody diarrhea and abdominal cramps at very low infective doses and has been fatal to children, the elderly and others with underdeveloped immunity (Gage 2001). In the 1986 Health Effects Criteria for Fresh Recreational Waters, the US Environmental Protection Agency designated *E. coli* as an indicator organism due to the strong correlation between *E. coli* levels and the risk of swimming related gastroenteritis (US Environmental Protection Agency 1986). *E. coli* and other pathogens are common in the intestines of cattle and other warm blooded animals. As such, an understanding of the sources that are contributing to fecal contamination is necessary to allow researchers and managers to target sites for remediation efforts with greater precision, and provide more detailed information about the effectiveness of ongoing remediation (Leung *et al.* 2003).

Unfortunately, traditional methods used to quantify bacterial pollution, such as membrane filtration (Dufour 1984) give no indication of the source of contamination or where management practices should be targeted (Dombek *et al.* 2000). As a result, several methods have been developed to identify sources

of fecal pollution, including antibiotic resistance analysis (ARA), pulsed field gel electrophoresis (PFGE), denaturing gradient gel electrophoresis (DGGE), ribotyping and Rep-PCR (Dombek *et al.* 2000; Whitlock *et al.* 2002). Antibiotic resistance analysis (ARA) requires less processing time at subsequently lower costs, which enables researchers to greatly increase the size of the source library (Whitlock *et al.* 2002; Choi *et al.* 2003). Molecular libraries are often much smaller than ARA libraries, ranging from 91 to 154 isolates compared with nearly 2500 isolates in some ARA databases (Dombek *et al.* 2000; Holloway 2001; Whitlock *et al.* 2002; Choi *et al.* 2003). A disadvantage of ARA is that it can be effectively applied only if bacteria have been exposed to antibiotics prior to application of the technique. In addition, wildlife residing in and around agricultural lands may develop the same antibiotic resistance patterns as many domestic animals (Meays *et al.* 2004). Although they are more costly and time-consuming, DNA fingerprinting techniques such as PFGE, ribotyping and Rep-PCR are reported to be an effective means with which to track the source of microbial pollution, without many of the complications of antibiotic resistance analyses (Rademaker and DeBruijn 1998; Carson *et al.* 2002; Whitlock *et al.* 2002; Lu *et al.* 2004). Although Meays *et al.* (2004) concluded that the choice of a particular source tracking method was both investigator and project specific; Carson *et al.* (2002) found that Rep-PCR proved to be a more effective technique than ribotyping. The Rep-PCR method of biological source tracking (BST) uses the Polymerase Chain Reaction (PCR) to produce unique DNA banding patterns (fingerprints) that are strain-specific. By comparing fingerprints of sample *E. coli*

to fingerprints of known animal sources, the source of the bacterial pollution may be identified.

In the Finger Lakes region of western New York, wildlife resides in relatively high densities on watersheds dominated by dairy and cattle farms, and as a result, the sources of fecal contamination are called into question. Source tracking may provide insight on which animal sources are contributing to fecal contamination in these lakes and streams.

The watershed of Conesus Lake, the smallest and westernmost of the Finger Lakes provides a microcosm of the Finger Lakes region. Upland areas of the Conesus Lake watersheds are dominated by agricultural practices, particularly dairy and cash crop farms, which provide habitat and food for geese and deer. Lakeside homes crowd the near-shore areas of Conesus Lake watersheds with the lake itself heavily used for boating and fishing. The nearly 9,800 residents of the watershed could contribute to fecal pollution of the watershed's small streams by way of leaking septic systems. Septic leaks were reported twice within the time of the study, although no significant increases in *E. coli* or total coliforms could be correlated with these reports (Davin, R. personal communication; Simon, R. personal communication). Nonetheless, the potential exists for humans to contribute to fecal contamination in the Conesus Lake watershed.

High losses of nutrients and manure from farms in the sub-watersheds are also responsible for eutrophication of the lake and fecal contamination (Makarewicz *et al.* 2001). Increased macrophyte and algal biomass and beach

closings are common problems that are often blamed on farmers by the recreational users of the lake (Makarewicz *et al.* 2001). Implementation of Best Management Practices in the Conesus Lake watershed provides an extraordinary opportunity to use bacterial source tracking (BST) as an evaluative tool on selected agricultural sub-watersheds.

By using an experimental small-watershed approach and molecular bacterial source tracking techniques, we evaluated the following in several manipulated watersheds of Conesus Lake:

1. Identified and quantified the bacterial sources in managed watersheds.
2. Demonstrated the efficacy of the DNA fingerprinting biological source tracking method as a tool for tracking differences in *E. coli* distributions in watersheds with and without farms.
3. Evaluated the distribution of *E. coli* sources across seasons in the Conesus Lake watershed.
4. Evaluated the impact of hydrometeorological events on the distribution of *E. coli* sources in selected sub-watersheds.
5. Evaluated the impact of Best Management Practices on the distribution of *E. coli* sources in Conesus Lake watersheds.

Methods

Site Description:

Conesus Lake is the smallest of the Finger Lakes (181km^2) with nearly 9,800 people residing in its watershed. The lake is a source of drinking water for a population of nearly 15,000 and is a year-round recreation destination for many. Over half of the lake's drainage area is in dairy and/or cash crop farms and populations of geese and deer use habitat and food resources from agricultural areas (Conesus Lake: Watershed Management Plan 2003). Nearly twenty small and intermittent streams flow into the lake from upland areas with several running through these agricultural lands.

Experimental Design:

Four sub-watersheds were chosen for this study. Graywood Gully, Long Point and Southwest Creek each had a farm within the watershed area, while North McMillan had no farm and was the most heavily forested (Figure 1). Agricultural Best Management Practices (BMP's) were applied to Graywood Gully and Southwest Creek watersheds. In Graywood Gully watershed, strip crops and cover crops were implemented, manure spreading was stopped, 15,000 feet of sub-surface drainage was installed, and heifers on the farm were fenced out from waterways to deny them contact with ponds or streams (Herendeen, N. 2004). Management practices in Southwest Creek watershed included a nutrient management plan, and construction of manure lagoons for storage. The lagoons made it possible to stop manure spreading during winter.

Long Point control stream had one farm in the watershed that milked cows until July 2003, but by July 2003, the farm had removed almost all of its cows (Ceronie, K. personal communication). Manure is still applied on this farm, but it is injected rather than spread (Herendeen, N. 2004). Long Point watershed was considered a control stream because the farm within the watershed was not managed or manipulated in any way. North McMillan watershed was considered a control stream because it had no farm within its watershed boundary.

Isolation of E. coli *from fecal samples:*

E. coli was isolated from a total of thirty five fecal samples to provide source genetic fingerprints. Ten fecal samples each were collected aseptically from Canada geese (*Branta Canadensis*), humans (*Homo sapiens*) and Holstein cattle (*Bos taurus*), and five fecal samples were collected from white-tailed deer (*Odocoileus virginianus*). Goose fecal samples were taken from animals at Marketplace Mall in Henrietta, NY. Holsteins were sampled from Maxwell's farm in Geneseo, NY and human fecal samples were collected from ten volunteers from the Conesus Lake watershed. Feces from deer were collected at Mendon Ponds Park in Rochester, NY.

A slurry was formed by placing approximately 1 gram of fecal material from the sterile centrifuge tube in approximately 10 ml of sterile phosphate dilution buffer. Serial dilutions were made from the slurry and the *E. coli* isolates obtained by membrane filtration (Millipore 1991). After filtration, both the filter apparatus and test tube containing the dilution were rinsed with sterile dilution buffer and passed through the filter. The membrane filter was aseptically

transferred to a padded plate containing mColiBlue24 broth and incubated for 24 hours at 37°C. The mColiBlue24 broth produces blue colonies based on the reaction of the *E. coli* β-Glucuronidase with a chromogen (BCIG) in the broth (US Environmental Protection Agency 1999). Between each filtration, the filter apparatus was sterilized with boiling water.

Isolation of E. coli *from water samples/PCR methodology:*

E. coli colonies isolated from water samples at four watersheds throughout the winter and spring of 2003 and 2004 were kindly provided by Robert Simon at the State University of New York at Geneseo using membrane filtration (Dufour 1984). From the mColiBlue24 plates, single blue colonies were picked off, streaked onto Eosin Methylene Blue (EMB) agar and incubated at 37°C for 24-48 hours. After incubation, colonies that showed a metallic green sheen on the EMB were confirmed to be *E. coli* and placed in 5ml Luria Bertani (LB) broth for 24 hours at 37°C. Glycerol stocks were made by adding 400µl of a whole cell suspension to 100µl of 50% glycerol to form a 10% total solution of glycerol. *E. coli* stocks were frozen at -80°C (Winfrey *et al.* 1997).

E. coli isolates undergoing PCR analysis were taken from glycerol stock and resuspended in Luria Bertani broth for 24 hours at 37°C. From LB broth, 1ml of cells were centrifuged for 5-7 minutes to form a cell pellet. The pellet was resuspended with 1ml of a 1M NaCl solution and centrifuged again. The final cell pellet was resuspended in 100µl of sterile ddH_2O. 2 µl of this cell suspension was used as template DNA in the PCR reaction.

Each PCR reaction was prepared by adding 2 µl of template DNA to 1 µl of a 10mM concentration of dNTP's, 2µl of BOX A1R primer, 0.5µl Taq DNA polymerase, 5µl 10X PCR buffer and 14.5µl dH_2O for a total volume of 25µl PCR reaction mix. 25µl of mineral oil was added to the top of the reaction mix in order to retain the heat from the thermocycler. The BOX A1R primer targets inverted repeats of 154 bp found throughout the *E. coli* genome. Both the forward and reverse primers are the same: 5'-CTACGGCAAGGCGACGCTGACG-3' (Rademaker and DeBruijn 1998).

PCR reactions were carried out with a Perkin Elmer 480 DNA Thermal Cycler beginning with an incubation step at 95°C for 7 minutes and 35 cycles of 94°C for 1 minute, 53°C for 1 minute and 65°C for 8 minutes. The cycle was terminated with an extension step of 65°C for 16 minutes (Rademaker and DeBruijn 1998).

Samples from PCR reactions were run in a 1.5% agarose gel for 8-10 hours, at approximately 55-70 volts and stained with ethidium bromide (Winfrey *et al.* 1997). A 1Kb standard ladder (Novagen) was placed in the first and last lanes of the gel (Rademaker and DeBruijn 1998).

Digital Imaging and Statistics:

Lanes and bands from gel images, taken with a Kodak Capture IS 440 imager (Eastman Kodak, Rochester, NY) were assigned automatically using Kodak's 1D software with a band finding sensitivity of "0" or "–1". Appendix B describes in detail the method for assigning bands and calculating their

molecular weights and a typical gel is presented in Figure 2. The fingerprint pattern from each of the genetic source groups was compared to sample fingerprints by a Jaccard similarity coefficient (SPSS v.11.5 for Windows). Jackknife analysis demonstrated whether isolates would be assigned to the correct animal source (Dombek *et al.* 2000). Briefly, each isolate was manually assigned to the correct source group and one isolate was removed from the database. The removed isolate was compared to all other isolates in the source database. The removed isolate was then placed in the source group with the highest similarity to the removed isolate and the percentage of isolates that were correctly assigned to each source group was calculated. A similar process was used to assign an unknown isolate to a source group. In addition, discriminant analysis (SPSS version 11.5) provided a quantitative statistical approach to measure whether the multiple DNA bands are able to predict, or 'discriminate' the four known source groups (Leung *et al.* 2003). Discriminant analysis generates linear functions using the *E. coli* DNA bands that are input as variables. The Wilk's Lambda test statistic produces an F-value, which is calculated to test the significance of the functions. A lower Wilk's test statistic represents that a given function is more significant. These functions are used by SPSS to separate isolates into groups and produce a cluster plot (Figure 3) (Fisher and Belle 1993; US Environmental Protection Agency 2004).

Results

Sampling summary and reliability of source genetic library

A total of 123 *E. coli* isolates from four animal sources (cows, geese, humans, deer) were collected and served as the genetic source group library. Genetic fingerprints were obtained from individual *E. coli* isolates using PCR and electrophoresis (Figure 2) and each band in the fingerprint was scored based on its presence (1) or absence (0). A binary source group database of 0's and 1's was established using the fingerprints from 123 source isolates. Jackknife analysis, using maximum percent similarity, tested the ability of the source group database to correctly identify fingerprints from a particular source group. Humans were correctly identified 88% of the time; deer 77% of the time, geese 67% of the time and cows 56% of the time (Table 2). Cows were most often misidentified as geese (25%), but were also misidentified as deer (14%) and humans (6%). This trend was reciprocal for geese in that geese were misidentified as cows 26% of the time (Table 2). A similar result was evident using average group similarity (Dombek, *et al.* 2000). For example, humans were correctly assigned 85% of the time and deer 91% of the time. This result displays the ability of the source group database to correctly classify fingerprints, lending confidence that unknown samples will be correctly assigned as well.

The canonical discriminant plot indicates that bands created by BOX-PCR generate significant functions for discriminating the four source groups (Figure 3). Human, goose and deer isolates form three distinct clusters; however, cows

overlap into both the deer and goose clusters in the two-function plot (Figure 3). Canonical discriminant functions analysis verified the ability of genetic fingerprints to correctly predict the animal source group. Using the genetic source group database, a three-function model was generated with Wilk's lambda values of 0.061, 0.276 and 0.608 for functions 1-3, 2-3 and 3, respectively. An average of 90.2% of the original source group isolates were correctly predicted using these three functions (Table 3).

Unknown isolates (n=152) were collected from stream water in four sub-watersheds of Conesus Lake between the months of December 2003 and June 2004 for identification. The average similarity of a given unknown isolate to a source isolate was 55%, but ranged as high as 76% for some isolates.

Spatial and temporal trends in E. coli *source distributions*

To determine spatial and temporal changes in *E. coli* distribution in Conesus Lake sub-watersheds, unknown isolates were identified by source and the relative percentage of isolates in each animal source group was calculated. In general, the distribution of *E. coli* sources is similar across sub-watersheds with geese being the dominant source in each of the sub-watersheds (44.7% to 73.7% of the total sources), followed by cows (10.5% to 21.1%), deer (10.5% to 18.4%) humans (5.3% to 12.9%), and unidentified isolates (0.0% to 11.8%) (Table 4). Geese dominated winter distributions of *E. coli* sources in Graywood (69.2%), Southwest (66.7%) and North McMillan Creek (60.0%) (Table 5). Just three out of twenty-six *E. coli* isolates were identified as cows in Graywood, one

of six in Southwest Creek and only two of ten in North McMillan Creek. No isolates were identified as deer in Southwest Creek and no isolates were identified as humans in North McMillan Creek during winter 2003.

The largest percentages of *E. coli* isolates were identified as geese in all watersheds during spring 2003 and spring 2004, except for Long Point in spring 2003. At Long Point, cows and deer made up the highest percentage of isolates (31.2% in each source group) while just 25% of the isolates were identified as geese. In spring 2004, however, the Long Point watershed experienced a dramatic increase in the percentages of *E. coli* isolates identified as geese (from 25.0% to 59.1%). The percentage of isolates identified as cows changed in all of the sub-watersheds that were sampled between spring 2003 and spring 2004 (Table 6). In the Graywood Gully watershed *E. coli* isolates identified as cows dropped from 31.2% to 10.0%, in Southwest Creek from 18.2% to 5.9%, in Long Point from 31.2% to 13.6% and in North McMillan from 20.0% to 0.0% between 2003 and 2004. The number of isolates identified as humans ranged from 6.3% and 20.0% and those identified as deer ranged from 0.0% to 31.3% of the total isolates identified in the four sub-watersheds between spring 2003 and spring 2004.

Influence of hydrometeorological events on E. coli *distributions*

E. coli was isolated from both non-event and event samples in order to comprehend the effect of hydrometeorological events on *E. coli* source distributions. All of the sub-watersheds with farms within their boundaries showed

an increase in the number and percentages of isolates identified as cows during hydrometeorological events (Table 7). North McMillan is forested with no farm within the watershed, and during hydrometeorological events, no isolates (0.0%) were identified as cows while two isolates (20.0%) were identified as cows during non-event periods. Isolates that were identified as geese, deer and humans revealed no observable trend with regard to distributions during event and non-event conditions.

Discussion

Demonstration of the efficacy of the DNA fingerprinting method as a tool for tracking E. coli *distributions*

Previous literature has reported varying results using BOX-PCR methods of source identification. Dombek (2000) reported clear distinctions between source groups with little to no overlap in the Jackknife analysis using a library of 154 isolates; however, Holloway (2001) found fingerprints indistinguishable between source groups, citing a small sample size (91 isolates) as the problem. Using a somewhat larger source genetic database than Holloway (123 isolates), jackknife and discriminant analysis indicated that our technique was successful, separating the genetic fingerprints of the four major source groups (cows, geese deer and humans). However, the banding patterns for geese and cows had the greatest within group variability of all the sources (Table 2). This could confound results, had the relationship not been reciprocal. The percentage of misidentifications of geese and cows were equal, thus misidentifications of unknown *E. coli* isolates would not be skewed disproportionately toward either source group.

Identification and quantification of bacterial sources in managed watersheds

Building on previous research (Rademaker and DeBruijn 1998; Dombek *et al.* 2000; Holloway 2001 and Leung *et al.* 2003), this work sought not only to create a genetic source database, but also to use the library as an evaluative tool for tracking changes in Best Management Practices on experimental sub-

watersheds. Although others have used antibiotic resistance analysis for this purpose (Webster *et al.* 2004), the use of molecular fingerprinting as a tool of this kind represents a unique undertaking, which has yet to be attempted by any other research group.

The distribution of *E. coli* sources in the Conesus Lake watershed is shown in Figure 4. Since land use in the sub-watersheds was dominated by dairy and cash crop farms, it was expected that the majority of *E. coli* isolated would be identified as cows; however, an unexpectedly high percentage of isolates were identified as wildlife (geese and deer) throughout the year (Figure 4). The bacterial source tracking analysis is supported by traditional data from the Conesus Lake watershed. From 1997-2001, an average of 377 (59 to 863) sightings of geese were observed in January (Swift 1998). Swift estimates that a single adult Canada goose contributes 200 grams of droppings per day to the environment, equal to 11.8 to 172.6 kg of feces per month. Clearly, geese populate watersheds that contain farms, where habitat and food exists continually. Similarly, the deer population of Conesus Lake watershed was estimated at 2,092 in 2003 (Kirsch, A. personal communication). Bacterial source tracking analyses are supporting traditional population abundance analyses in suggesting that wildlife, especially geese and deer are contributing to the distribution of *E. coli* to sub-watersheds.

Others have reported comparable water quality degradation from wildlife. Whitman and Nevers (2003) indicated that gulls were a major source of fecal pollution in foreshore sands of 63rd Street beach on Lake Michigan. Choi *et al.*

(2003) showed that bird feces accounted for an average of 30% and as much as 67% of enterococci in the waters of Huntington State Beach, CA.

Evaluation of the impact of hydrometeorological events on the distribution of E. coli *sources in selected sub-watersheds*

Hydrometeorological events appear to influence sources of *E. coli* found in stream water draining agricultural watersheds (Table 7). For example, Long Point and Southwest Creek watersheds had three times the number of cow *E. coli* isolates during events when compared with non-event periods. This result confirms unpublished data in which elevated *E. coli* levels were observed with hydrometeorological events (Simon, R. personal communication). In addition, a segment analysis of events at Southwest Creek in which nutrient and bacterial sources were identified by sampling the stream in smaller and smaller segments reported the highest levels of coliforms (>16,000 CFU/100ml) and *E. coli* (30,000 CFU/100ml) at a site associated with an agricultural area (Makarewicz and Lewis 2002). All other sampling sites further downstream of the agricultural area ranged from just <103 to 500 CFU/100ml (Makarewicz and Lewis 2002). The segment analysis in 2002 supports the finding here that agricultural practices are contributing more to the distribution of *E. coli* in stream water during event periods.

Additionally, North McMillan, which has no farms in the watershed, would be expected to exhibit no shift in *E. coli* sources during event periods. As expected, a very low percentage of isolates were identified as cows, and both

were isolated from non-event samples, which shows a clear distinction between those watersheds with farms (Graywood Gully, Southwest Creek and Long Point) and those without farms (North McMillan). The fact that North McMillan had such a low percentage of *E. coli* isolates identified as cows lends further confidence that this genetic fingerprinting technique is accurately identifying unknown isolates by their respective sources.

Evaluation of the temporal trends in E. coli source distribution and the impact of Best Management Practices on source distributions in Conesus Lake watersheds

Management practices on farms in experimental watersheds were expected to reduce percentages of *E. coli* isolates identified as cows. This, in fact, did occur in Graywood Gully and the Southwest Creek experimental sub-watersheds (Table 6), although a lower percentage of isolates were identified as cows in the winter than in the spring of 2003. Midwinter spreading of manure is limited by cold and harsh conditions and is also not considered to be a good practice. Manure spread over ice and snow on frozen fields has a greater potential to be washed off during rain or snowmelt events. By the spring 2003 sampling season, the ground had thawed and the farms began routine manure spreading practices leading to an increase in cow *E. coli* isolates as identified by genetic fingerprinting. As a result, during the following year, Best Management Practices were begun. One of these management strategies was to reduce manure spreading on the steep slopes of experimental sub-watersheds in Conesus Lake. By 2004, the percentage of isolates identified as cows had

dropped off in experimental watersheds (from 31.3% to 10.0% in Graywood Gully and from 18.2% to 5.9% in Southwest Creek).

The Long Point control watershed also exhibited a decrease in isolates identified as cows between spring 2003 (31%) and spring 2004 (14%). In the spring 2003 sampling season (Ceronie, K. personal communication) the Long Point farm had been milking approximately 50 cows. By July 2003, the farm on Long Point had removed nearly all of its cows and the reduction in cow isolates identified by genetic fingerprinting may be indicative of this fact. The isolates that were identified as cows during spring 2004 can be attributed to continued manure injections at the Long Point farm.

The results from Graywood Gully, Southwest Creek and Long Point suggest that management practices are effectively reducing *E. coli* coming from cows; however, one of the limitations of the genetic fingerprinting method is the potential for misidentifications of unknown isolates to the wrong animal source group. For example, the discriminant functions plot (Figure 3) visually represents this potential for animal source groups to overlap. In order to further substantiate the results that *E. coli* from cows are being reduced in experimental watersheds a conservative approach was used to assign isolates to the cow group. This was done by assigning an unknown isolate to the cow source group using $\geq$ 50% similarity, regardless of whether that unknown was more similar to another animal source. Even when excluding the potential to misidentify cow isolates as another animal source, there is a decline in the number and percentage of isolates that were identified as cows between spring 2003 (50% in Graywood,

42% in Southwest Creek) and spring 2004 (35% in Graywood, 24% in Southwest Creek) in experimental watersheds (Figure 5). This decline appears to demonstrate the effects of the Best Management Practices and certainly lends further support that management practices are leading to reductions in *E. coli* coming from cows. It is important to note, however, that these data represent changes in percentages of isolates from a small sample size and are not subject to statistical tests for significance.

In conclusion, BOX A1R-derived genetic fingerprints are useful in identifying *E. coli* isolates by source and may be useful in assessing the impact of implemented BMP's on Conesus Lake sub-watersheds. These techniques, however, require expensive equipment and well-trained technicians. Because of this, the genetic source identification technique is not recommended for small budget municipalities. As an alternative, antibiotic resistance analysis may provide an easier and more economical approach to source identification (Meays *et al.* 2004; Whitlock et al. 2002).

Literature Cited

1. Carson, C. A., Shear, B. L., Ellersieck, M. R., Schnell, J. 2002. Comparison of ribotyping and repetitive extragenic palindromic-PCR for identification of fecal *Escherichia coli* from humans and animals. Applied and Environmental Microbiology, 69 (3): 1836-1839.
2. Ceronie, K. New York State Farm Service Agency. 129 Main St. P.O. Box 196, Leicester, NY 14481. 585-382-3221. Personal Communication.
3. Choi, S., Chu, W., Brown, J., Becker, S. J., Harwood, V. J., Jiang, S. C. 2003. Application of enterococci antibiotic resistance patterns for contamination source identification at Huntington Beach, California. Marine Pollution Bulletin, 46: 784-755.
4. Conesus Lake: Watershed Management Plan, 2003. Livingston County Planning Department. Geneseo, NY 14454.
5. Davin, R. Senior Sanitarian for Livingston County Department of Health and Conesus Lake Watershed Inspector. Center for Environmental Health. 2 Murray Hill Drive. Mount Morris, New York 14510. (585)-243-7280. Personal Communication.
6. Dombek, P. E., Johnson, L. K. Zimmerley, S. T., Sadowsky, M. J. 2000. Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. Applied and Environmental Microbiology, June 2000, 2572-2577.

7. Dufour, A. P. 1984. Health Effects Criteria for Fresh Recreational Waters. EPA-600/1-84-004, U.S. Environmental Protection Agency, Cincinnati, Ohio.
8. Environmental Protection Agency. 5/10/2004. EPA Statistical Primer. http://www.epa.gov/bioindicators/primer/dfa.html.
9. Fisher, L. D. and Van Belle, G. 1993. Biostatistics: A Methodology for the Health Sciences. John Wiley and Sons, Inc. New York, NY.
10. Gage, R., Crielly, A., Baysinger, M., Chernak, E., Herbert, A., Johnson-Entsuah, A., Fraser, G., Rinehardt, C., Solomon, M., Withers, G., Berman, R., Moll, M., Rankin, J., Carroll, J., Ettinger, M., Henderson, S., Mismas, M., Patel, D., Reed, T., Smith, E., Wozniak, J., Toney, D., Pearson, J., Hofmann, J., Grendon, J., Kobayashi, J. 2001. Outbreaks of *Escherichia coli* O157:H7 Infections Among Children Associated with Farm Visits-Pennsylvania and Washington, 2000. Centers for Disease Control. Morbidity and Mortality Weekly Report, 50 (15): 293-297.
11. Herendeen, N. 2004. Best Management Practices Summary Submitted to Cornell Cooperative Extension NWNY Team. 420 East Main St. Batavia, NY 14020.
12. Holloway, P. 2001. Tracing the source of *E. coli* fecal contamination of water using rep-PCR. Manitoba Livestock Manure Management Initiative Project: 00-02-08.
13. Kingsley, K. Applied Maths, Inc. 512-482-9700. Personal Communication.

14. Kirsch, A. Senior Wildlife Biologist. New York State Department of Environmental Conservation. Bureau of Wildlife. 6274 E. Avon-Lima Rd. Avon, NY 14414. (585) 226-5461. Personal Communication.

15. Leung, K. T., Mackereth, R., Tien, Y., Topp, E. 2003. A comparison of AFLP and ERIC-PCR analyses for discriminating *Escherichia coli* from cattle, pig and human sources. Microbiology Ecology, 47: 111-119.

16. Lu, L., Hume, M. E., Sternes, K. L., Pillai, S. D. 2004. Genetic diversity of *Escherichia coli* isolates in irrigation water and associated sediments: implications for source tracking. Water Research, 38: 3899-3908.

17. Makarewicz, J. C., Bosch, I., Lewis, T. W. 2001. Soil and nutrient loss from selected subwatersheds of Conesus Lake. Finger Lakes-Lake Ontario Watershed Protection Alliance. Livingston County Planning Department. Geneseo, NY.

18. Makarewicz, J. C. and Lewis, T. 2002. Stressed Stream Analysis of North Gully and Southwest Creeks In the Conesus Lake Watershed. Center for Applied Aquatic Science and Aquaculture. Department of Environmental Science and Biology, SUNY Brockport. Brockport, NY 14420.

19. Meays, C. L., Broersma, K., Nordin, R. Mazumder, A. 2004. Source tracking fecal bacteria in water: a critical review of current methods. Journal of Environmental Management, 73 (2004): 71-79.

20. Millipore Corporation. 1991. Microbiological Analysis of Water and Wastewater. Millipore Corporation. Bedford, MA. 01730.

21. Rademaker, J. L. W. and DeBruijn, F. J. 1998. Characterization and classification of microbes by Rep-PCR genomic fingerprinting and computer-assisted pattern analysis. Applied and Environmental Microbiology, 64: 2096-2104.

22. Simon, R. State University of New York at Geneseo. Geneseo, NY 14454. Personal Communication.

23. Swift, B. L. 1998. Impacts of Canada Geese on Water Quality and Public Health: A Preliminary Assessment. New York State Department of Environmental Conservation Bureau of Wildlife.

24. US Environmental Protection Agency. 1999. National Primary Drinking Water Regulations: Methods for Microbes, Lead and Magnesium. Federal Register, 64 (9): 2537-2544

25. US Environmental Protection Agency. 1986. Ambient Water Quality Criteria for Bacteria. EPA 440/5-84-002. EPA Office of Water Regulations and Standards. Criteria and Standards Division. Washington, D. C. 20460.

26. Webster, L. F., Thompson, B. C., Fulton, M. H., Chestnut, D. E., Van Dolah, R. F., Leight, A. K., Scott, G. I. 2004. Identification of sources of *Escherichia coli* in South Carolina estuaries using antibiotic resistance analysis. Journal of Experimental Marine Biology and Ecology, 298 (2): 179-195.

27. Whitlock, J. E., Jones, D. T., Harwood, V. J. 2002. Identification of the sources of fecal coliforms in an urban watershed using antibiotic resistance analysis. Water Research, 36: 4273-4282.

28. Whitman, R. L. and Nevers, M. B. 2003. Foreshore Sand as a Source of *Escherichia coli* in Nearshore Water of a Lake Michigan Beach. Applied and Environmental Microbiology, 69 (9): 5555-5562.

29. Winfrey, M. R., Rott, M. A., Wortman, A. T. 1997. Unraveling DNA: Molecular Biology for the Laboratory. Prentice Hall. Upper Saddle River, NJ.

Watershed Name	*Watershed Type*	*Range of Sampling Dates*	*Number of Sampling Dates*	*Event Isolates*	*Non-Event Isolates*	*Total Isolates*
Graywood Gully	Experimental	12/16/2002 to 4/18/2004	11	24	38	62
Southwest Creek	Experimental	12/31/2002 to 6/2/2004	9	13	22	35
Long Point	Control	3/16/2003 to 6/1/2004	13	21	15	36
North McMillan	Control	2/4/2003 to 6/2/2003	5	9	10	19
						152

Table 1: Summary of watershed *E. coli* sampling. Four sub-watersheds were sampled: two watersheds contained farms with implemented best management practices (experimental watersheds) while two watersheds did not (control watersheds). Each sub-watershed is listed along with the range of sampling dates, number of sampling dates and breakdown of sample sizes of *E. coli* isolated from event and non-event samples. Total isolates for each watershed are depicted in the far right column.

	Cows	Geese	Deer	Humans
Cows	**55.5%**	*25.0%*	*13.8%*	*5.5%*
Geese	*25.6%*	**66.6%**	*5.1%*	*2.6%*
Deer	*18.2%*	*4.5%*	**77.3%**	*0.0%*
Humans	*0.0%*	*0.0%*	*11.5%*	**88.5%**

Table 2: Jackknife analysis demonstrates the percentage of *E. coli* source isolates that were correctly assigned to each source group. Highlighted cells represent the percentage of *E. coli* isolates assigned to the correct source group. Cows were correctly assigned 55.5% of the time, geese 66.6%, deer 77.3% and humans 88.5% of the time. Cows were most often misidentified as geese (25.0%) and geese were most often misidentified as cows (25.6%).

Predicted Group Membership	Cows	Geese	Deer	Humans
Cows	**86.1%**	8.3%	5.6%	0.0%
Geese	5.1%	**89.7%**	2.6%	2.6%
Deer	0.0%	13.6%	**86.4%**	0.0%
Humans	0.0%	0.0%	0.0%	**100.0%**

90.2% of original grouped cases correctly classified.

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	Degrees of Freedom	Significance (p)
1 through 3	0.0609	272.8405	135	0.0000
2 through 3	0.2761	125.4650	88	0.0054
3	0.6083	48.4591	43	0.2622

Table 3: Wilk's Lambda discriminant functions analysis of *E. coli* isolates. Analysis demonstrated that the correct source group is assigned an average of 90.2% of the time. Analysis also indicated that functions 1 and 2 separated fingerprints into groups with statistical significance ($p \leq 0.005$).

	Humans	Geese	Cows	Deer	Unidentified	Total Isolates
Graywood Gully	12.9% (8)	53.2% (33)	16.1% (10)	14.5% (9)	3.2% (2)	62
Southwest Creek	11.8% (4)	52.9% (18)	11.8% (4)	11.8% (4)	11.8% (4)	34
Long Point	7.9% (3)	44.7% (17)	21.1% (8)	18.4% (7)	7.9% (3)	38
North McMillan	5.3% (1)	73.7% (14)	10.5% (2)	10.5% (2)	0.0% (0)	19

Table 4: Distribution of *E. coli* sources in Conesus Lake sub-watersheds. The percentage of unknown isolates assigned to each source group is listed with the number of isolates assigned to each group in parentheses. Total isolates = the total number of *E. coli* isolates analyzed for each watershed.

	Humans	Geese	Cows	Deer	Unidentified	Total Isolates
Graywood Gully	11.5% (3)	69.2% (18)	11.5% (3)	3.8% (1)	3.8% (1)	26
Southwest Creek	16.7% (1)	66.7%(4)	16.7% (1)	0.0% (0)	0.0% (0)	6
North McMillan	0.0% (0)	60.0% (6)	20.0% (2)	20.0% (2)	0.0% (0)	10

Table 5: Distribution of *E. coli* sources in Conesus Lake sub-watersheds: Winter 2003. The percentage of unknown isolates assigned to each source group from winter 2003 is listed with the number of isolates assigned to each group in parentheses. Total isolates = the total number of *E. coli* isolates analyzed for each watershed. Source groups run left to right and sub-watersheds run top to bottom.

	Humans	Geese	Cows	Deer	Unidentified	Total Isolates
Graywood Gully Spring 2003	6.3% (1)	37.5% (6)	31.3% (5)	25.0% (4)	0.0% (0)	16
Graywood Gully Spring 2004	20.0% (4)	45.0% (9)	10.0% (2)	20.0% (4)	5.0% (1)	20
Southwest Creek Spring 2003	18.2% (2)	54.5% (6)	18.2% (2)	0.0% (0)	9.1% (1)	11
Southwest Creek Spring 2004	5.9% (1)	47.1% (8)	5.9% (1)	23.5% (4)	17.6% (3)	17
Long Point Spring 2003	6.3% (1)	25.0% (4)	31.3% (5)	31.3% (5)	6.3% (1)	16
Long Point Spring 2004	9.1% (2)	59.1% (13)	13.6% (3)	9.1% (2)	9.1% (2)	22

Table 6: Distribution of *E. coli* sources in Conesus Lake sub-watersheds: Spring 2003 and Spring 2004. The percentage of unknown isolates assigned to each source group is listed with the number of isolates assigned to each group in parentheses. Total isolates = the total number of *E. coli* isolates analyzed for each watershed. Source groups run left to right and sub-watersheds run top to bottom.

	Humans	Geese	Cows	Deer	Unidentified	Total Isolates
Graywood Gully Non-Event	14.3% (2)	35.7% (5)	28.6% (4)	21.4% (3)	0.0% (0)	14
	3	10	3	5	1	
Graywood Gully Event	13.6% (3)	45.5% (10)	13.6% (3)	22.7% (5)	4.5% (1)	22
	1	10	0	2	4	
Southwest Creek Non-Event	5.9% (1)	58.5% (10)	0.0% (0)	11.8% (2)	23.5% (4)	17
	2	4	3	2	1	
Southwest Creek Event	16.7% (2)	33.3% (4)	25.0% (3)	16.7% (2)	8.3% (1)	12
	3	8	2	2	2	
Long Point Non-Event	17.6% (3)	47.1% (8)	11.8% (2)	11.8% (2)	11.8% (2)	17
	0	9	6	5	1	
Long Point Event	0.0% (0)	42.9% (9)	28.6% (6)	23.8% (5)	4.8% (1)	21
	0	6	2	2	0	
North McMillan Non-Event	0.0% (0)	60.0% (6)	20.0% (2)	20.0% (2)	0.0% (0)	10
	1	8	0	0	0	
North McMillan Event	11.1% (1)	88.9% (8)	0.0% (0)	0.0% (0)	0.0% (0)	9

Table 7: Distribution of *E. coli* sources during hydrometeorological events and non-event periods. The percentage of unknown isolates assigned to each source group from winter 2003 is listed with the number of isolates assigned to each group in parentheses. Total isolates = the total number of *E. coli* isolates analyzed for each watershed. Source groups run left to right and sub-watersheds run top to bottom (non-event followed by event conditions).

Figure 1: Four sub-watersheds were chosen for this study. Graywood Gully, Long Point and Southwest Creek each had a farm within the watershed area, while North McMillan had no farm and was the most heavily forested.

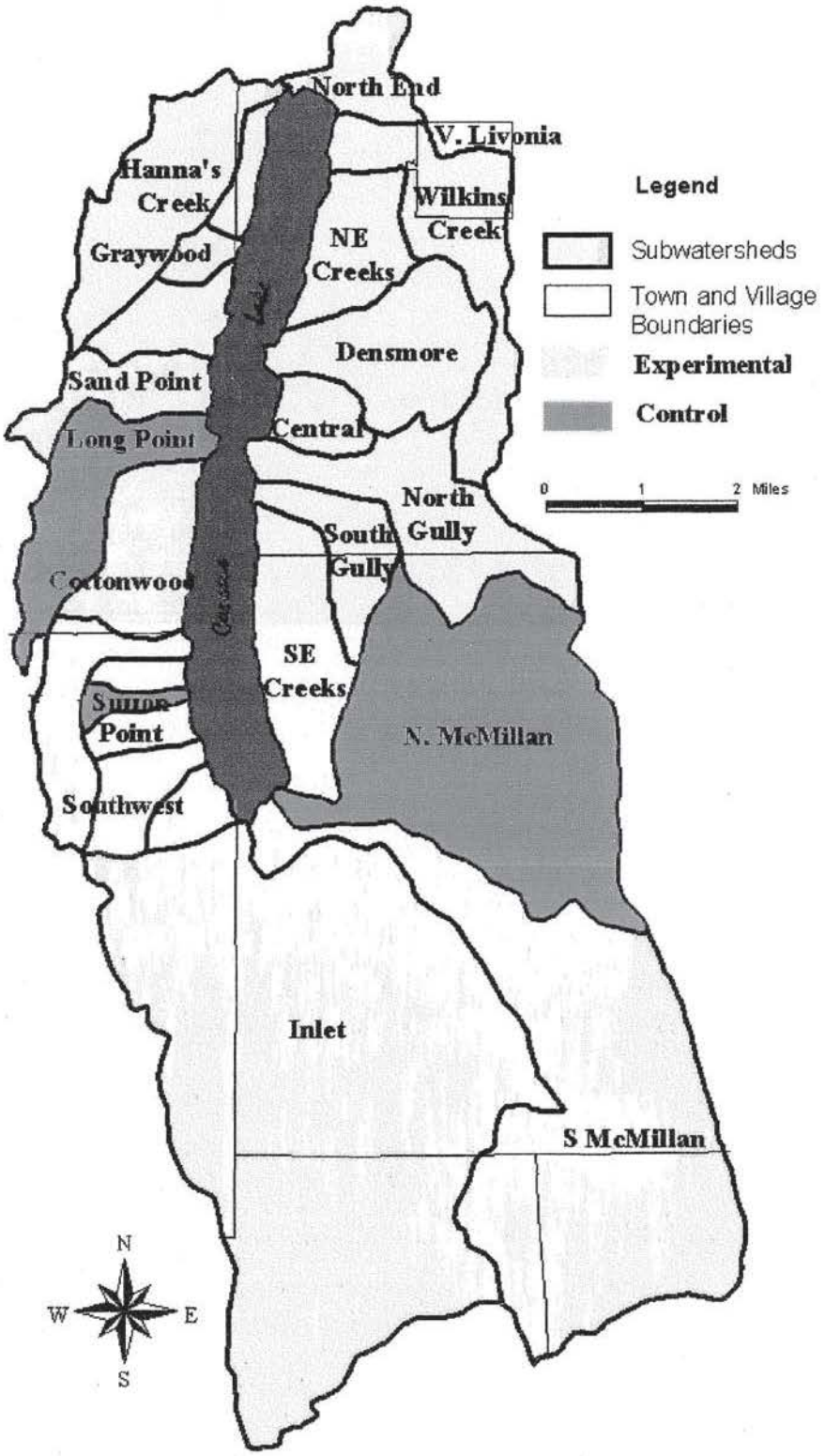


Figure 2: A typical gel is shown below. Lanes 1, 15 and 30 contain a 1Kb DNA ladder (Novagen). All other lanes contain BOX-derived *E. coli* fingerprints isolated from humans. A detailed explanation of the method for assigning bands to a binary matrix can be found in Appendix B: Gel Normalization Method.

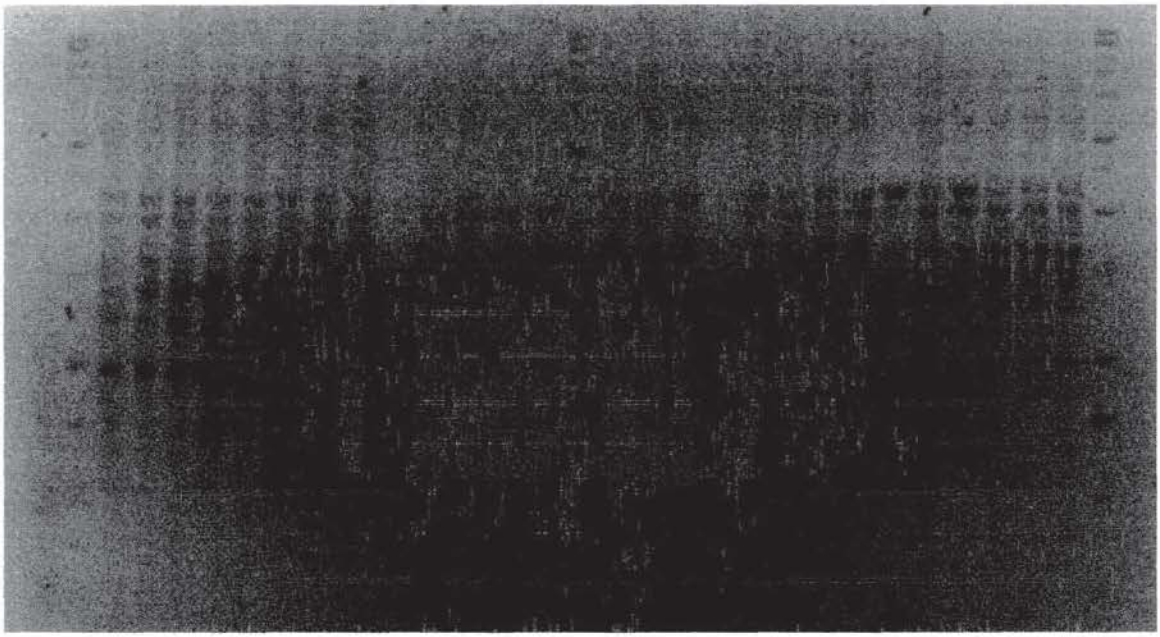


Figure 3: Canonical Discriminant Functions Analysis of animal source genetic fingerprints. Source groups form distinct clusters, with the cow source overlapping between deer and geese source isolates.

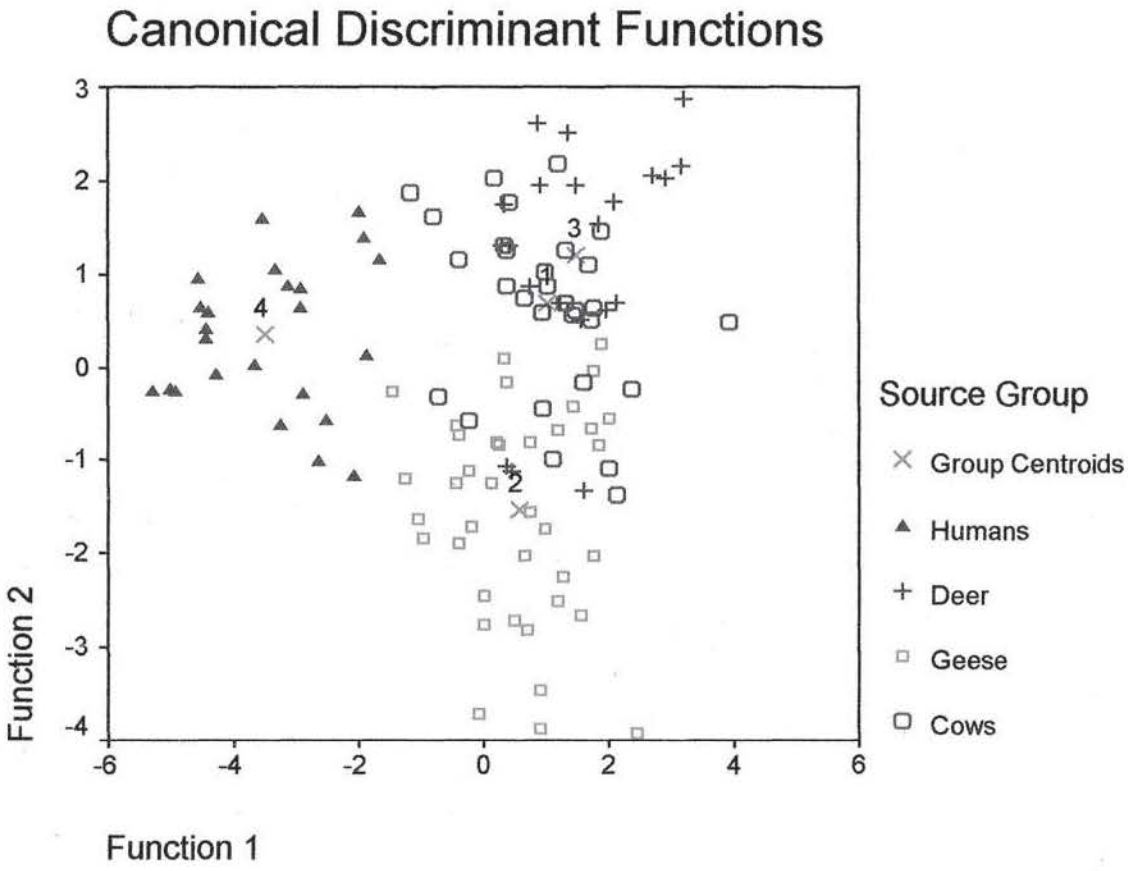


Figure 4: Distribution of *E. coli* sources in Conesus Lake sub-watersheds: Percentages above bars represent the relative percentage of isolates identified as a given animal source in each watershed.

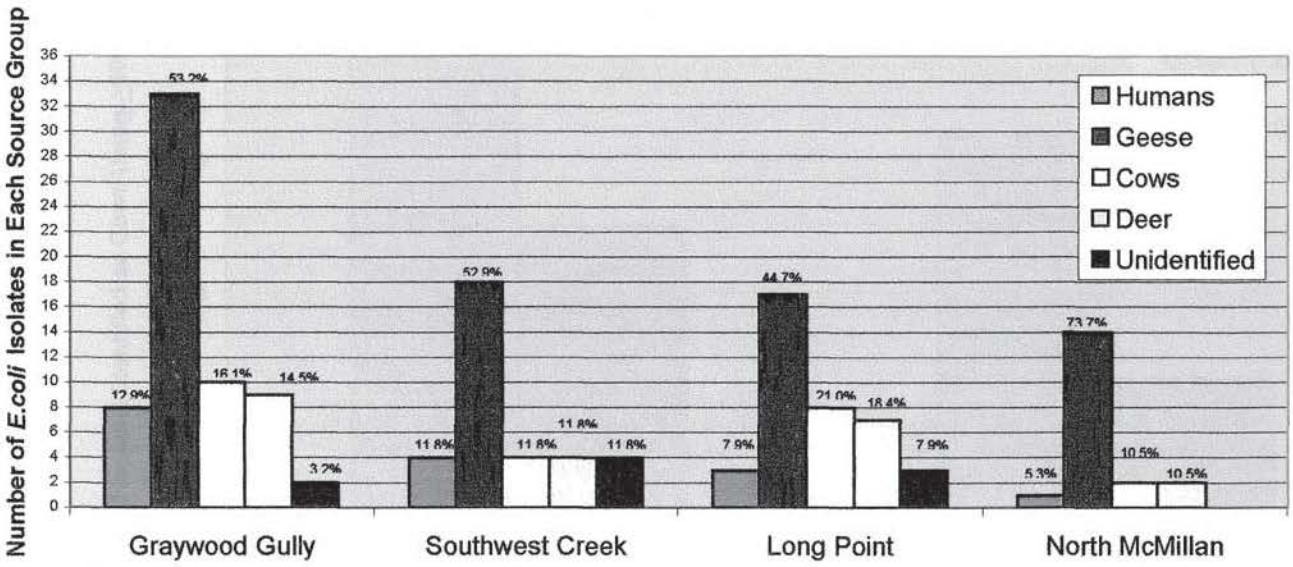
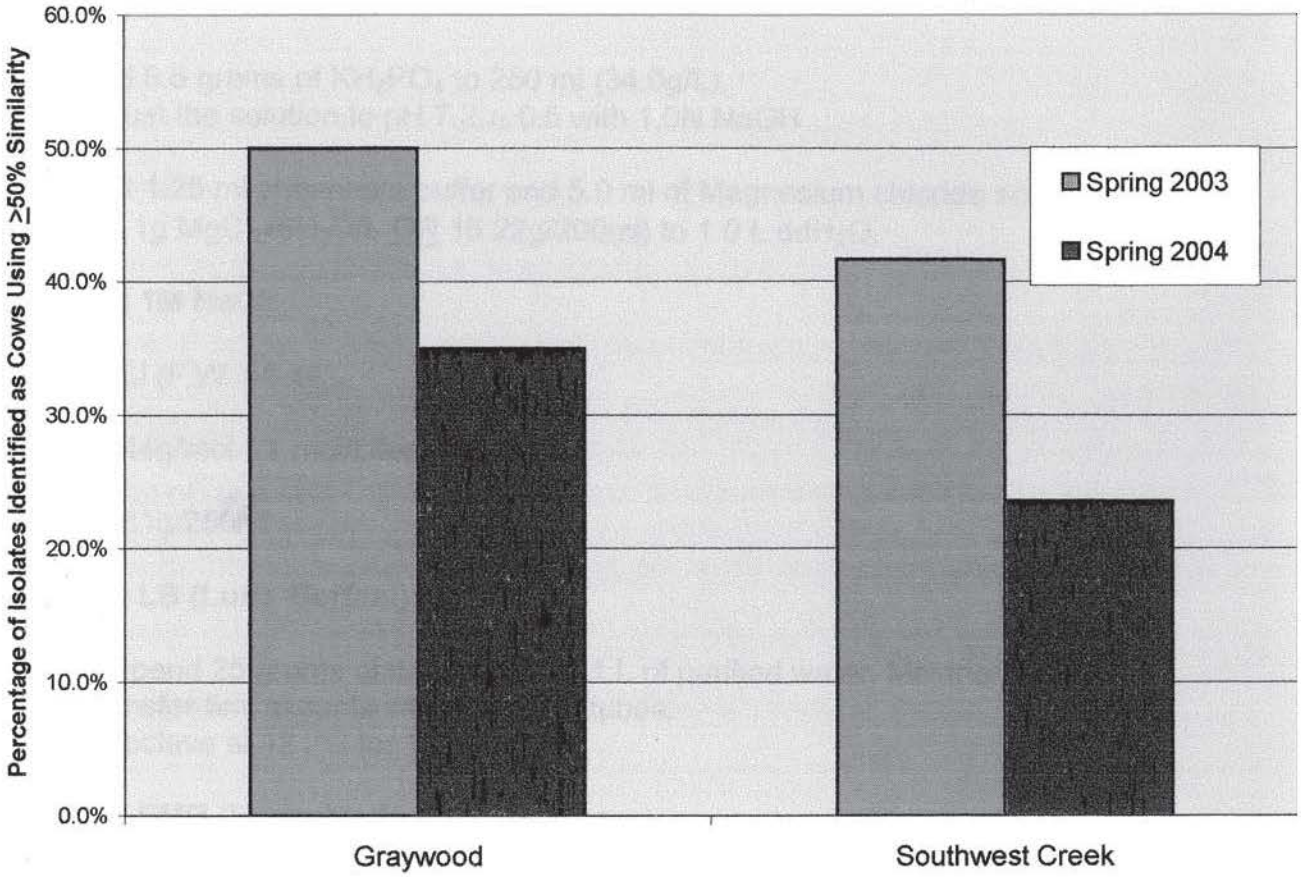


Figure 5: Change in Agricultural Source Contribution to Experimental Watersheds (2003-2004) Using ≥50% Similarity



Appendix A: Reagents List

Preparing Buffered Water:

Add 8.5 grams of KH_2PO_4 to 250 ml (34.0g/L).
Adjust the solution to pH 7.2 $+/-$ 0.5 with 1.0N NaOH

Add 1.25 ml phosphate buffer and 5.0 ml of Magnesium chloride solution (81.1g $MgCl_2 \bullet 6H_2O$/L **OR** 16.22g/200ml) to 1.0 L ddH_2O.

Preparing 1M NaCl:

NaCl (F.W. 58.44)

58.44g/mol * 1 mol/Liter = 58.44g/L
OR
14.61g/250ml

Preparing LB (Luria-Bertani):

Suspend 25 grams of the powder in 1 L of purified water. Mix thoroughly.
Transfer 5ml aliquots into clean test tubes.
Autoclave at 121°C for 15 minutes.

Preparing EMB (Eosin Methylene Blue) agar:

Suspend 36 grams of the powder in 1 L of purified water. Mix thoroughly.
Cover with Parafilm and heat in a microwave for 2-4 minutes or until completely dissolved.
Autoclave at 121°C for 15 minutes.
Pour approximately 20ml of hot (~65°C) culture into each plate and let cool. Store plates upside down in refrigerator.

Preparing 0.5X TAE buffer:

Add 12.5 ml 10X TAE to a 500 ml Erlenmeyer flask and dilute to 250ml with ddH_2O.
OR 25ml 10X TAE/500ml
OR 50 ml 10X TAE/1.0L

Preparing 1.5% agarose:

In a 500 ml bottle add 3.75g agarose to 250 ml 0.5X TAE and dissolve the agarose by microwaving until the agarose appears clear and gelatinous.
Allow agarose to cool and add 5-10µl stock Ethidium Bromide (10mg/ml).

Appendix B: Standard Operating Procedure for Bacterial Source Tracking with BOX A1R Primed PCR

Fecal material was collected from animal sources in 50 ml plastic centrifuge tubes. Samples were capped and stored on ice until processing. All samples were processed within 24 hours. Approximately 1 gram of fecal material was removed from the centrifuge tube and placed in a second sterile 50 ml tube containing approximately 10 ml of sterile phosphate dilution buffer (Appendix A). A sterile weighing spoon was used to break up the feces and the samples were inverted 15-20 times to ensure proper mixing. Serial dilutions were made out to 10^{-3} by adding 1ml of slurry to 9ml of sterile phosphate dilution buffer in a test tube. No solid fecal material was included in the serial dilutions. *E. coli* isolates were obtained by membrane filtration of all serial dilutions (Millipore 1991). After filtration, both the filter apparatus and test tube containing the dilution were rinsed with sterile dilution buffer and passed through the filter. The membrane filter was then aseptically transferred using alcohol flamed forceps to a padded plate containing mColiBlue24 broth and incubated for 24 hours at 37°C. Between each separate filtration, the filter apparatus was sterilized with boiling water.

After 24 hours incubation, single blue colonies were removed from mColiBlue24 plates with a sterile loop, streaked onto Eosin Methylene Blue (EMB) agar and incubated at 37°C for 24-48 hours. Colonies that showed a metallic green sheen on the EMB were confirmed to be *E. coli* and placed in 5ml Luria Bertani (LB) broth with a sterile loop for 24 hours at 37°C. Glycerol stocks

were made by adding 400µl of a whole cell suspension to 100µl of 50% glycerol to form a 10% total solution of glycerol. 50% glycerol was added to each 1.5 ml microcentrifuge tube first, followed by addition of cell suspension. *E. coli* stocks were frozen at -80°C (Winfrey *et al.* 1997).

E. coli isolates undergoing PCR analysis were taken from glycerol stock and resuspended in Luria Bertani broth for 24 hours at 37°C. From LB broth, 1ml of cells was centrifuged for 5-7 minutes to form a cell pellet. The pellet was resuspended with 1ml of a 1M NaCl solution and centrifuged again. The addition NaCl wash was repeated at least two times. The final cell pellet was resuspended by drawing the liquid up and down 15-20 times in 100µl of sterile ddH_2O. 2 µl of this cell suspension was used as template DNA in the PCR reaction.

Each PCR reaction was prepared separately by adding 14.5µl dH_2O, 5µl 10X PCR buffer, 2µl of BOX A1R primer, 1 µl of 10mM dNTP's, 2 µl of template DNA and 0.5µl Taq DNA polymerase for a total volume of 25µl PCR reaction mix. Each reaction was mixed thoroughly and 25µl of mineral oil was added to the top of the reaction mix. A master mix could be prepared with all reagents except the Taq, which could be mixed thoroughly into the reaction after the initial 7-minute incubation step. A master mix would eliminate pipetting errors and addition of Taq after the incubation step would prolong the fidelity of the enzyme.

PCR reactions were carried out with a Perkin Elmer 480 DNA Thermal Cycler beginning with an incubation step at 95°C for 7 minutes and 35 cycles of 94°C for 1 minute, 53°C for 1 minute and 65°C for 8 minutes. The cycle was

terminated with an extension step of 65°C for 16 minutes and soak at 4°C overnight (Rademaker and DeBruijn 1998).

12 µl of PCR product was added to 2.4 µl of 6X loading dye, loaded into a 1.5% agarose gel (Appendix A) and run for 8-10 hours at approximately 55-70 volts (Winfrey *et al.* 1997). 10 µl of ethidium bromide was added directly to melted agarose upon cooling and a 1Kb standard ladder (Novagen) was placed in the first and last lanes of the gel by adding 2 µl ladder, mixed with 4 ml ddH_2O and 1.2 µl loading buffer (Rademaker and DeBruijn 1998).

A detailed protocol for gel imaging and normalization across gels is described in Appendix C: Gel Normalization Method.

Appendix C: Gel Normalization Method

In order to designate whether two bands are the same, the total distance of the gel must first be assigned. This is done by cropping the gel image using the Kodak 1D software. Typically, the total gel distance runs from the first molecular weight marker to the last (Figure 2). Both lanes and bands are assigned automatically using the Kodak 1D software and the sensitivity of the band finding algorithm can be adjusted to account for noise in the gel image. Increasing the sensitivity will find more bands and decreasing the sensitivity will find fewer bands. After lanes and bands are assigned using Kodak 1D software, the molecular weight of each band is calculated by the software using the molecular weight markers as a reference. Using the total distance of the molecular weight markers, the sum of the range of the molecular weights is divided by the number of lanes that contain a molecular weight marker. The average range of the molecular weights is then divided by 100 to get 1% of the total distance of the gel, also called position tolerance (Kingsley, K., Personal Communication). GelCompar II software typically uses a 1% difference in band distance as the position tolerance for matching bands across lanes (Kingsley, K., Personal Communication). The calculated molecular weight ± 1% position tolerance is used to designate whether two bands are the same.

Using Excel, a standardized method for assigning presence or absence of bands across gels was developed. Molecular weights were inputted in a single column starting with 12,500 base pairs and ending with 300 base pairs in

increments of 150 base pairs. This list of molecular weights was used as a standard for assigning presence or absence of bands in each gel lane. If a particular lane contained a band on the list, a 1 was assigned. If the band was absent in that lane, a 0 was assigned. The table of 0's and 1's was then loaded into SPSS and the Jaccard Similarity Coefficient was measured for each lane.

Appendix D: Summer 2003 and 2004 Data

	Cows	Geese	Deer	Humans	Total Isolates
Graywood Gully	27.8% (5)	27.8% (5)	22.2% (4)	22.2% (4)	18
Southwest Creek	18.8% (3)	62.5% (10)	6.3% (1)	12.5% (2)	16

Distribution of *E. coli* isolates in experimental sub-watersheds: Summer 2003 and 2004. The percentage of unknown isolates assigned to each source group from summer 2003 and 2004 is listed with the number of isolates assigned to each group in parentheses. Total isolates = the total number of *E. coli* isolates analyzed for each watershed. Source groups run left to right and sub-watersheds run top to bottom. Samples were collected seven times from 7/1/2003 to 7/30/2003 and from 6/29/2004 to 7/20/2004. Geese and deer constitute 50% of the *E. coli* distribution in Graywood Gully experimental watershed during the summer. Furthermore, geese alone make up nearly two thirds of the *E. coli* source distribution in Southwest Creek experimental watershed. These data further indicate that geese and other wildlife contribute greatly to *E. coli* distributions in Conesus Lake sub-watersheds.